

Đề thi cuối kỳ

Môn: **Hệ Quản Trị Cơ Sở Dữ Liệu (503004)**

Ngành: Khoa Học Máy Tính – HK1 – 2013-2014

Thời gian làm bài: **120** phút - Mã đề: **01**

(Bài thi gồm **25** câu hỏi trắc nghiệm và **3** câu hỏi tự luận.)

Sinh viên được **tham khảo ghi chú** trong 2 tờ giấy A4.)

I. PHẦN TRẮC NGHIỆM

Sinh viên chọn **1** câu trả lời đúng nhất và trả lời vào **Phần trả lời** ở cuối đề thi. Nếu chọn câu (e) thì sinh viên cần trình bày đáp án khác so với đáp án ở các câu (a), (b), (c), và (d) và/hoặc giải thích lựa chọn (e) của mình.

Câu 1. Các câu lệnh về định nghĩa dữ liệu (*data definition statement*) do **ai** gọi đến hệ quản trị cơ sở dữ liệu?

- a. Người thiết kế cơ sở dữ liệu (*database designer*)
- b. Người sở hữu dữ liệu (*data owner*)
- c. Người quản trị cơ sở dữ liệu (*database administrator – DBA*)
- d. Các chương trình ứng dụng (*application program*) chạy trên hệ quản trị cơ sở dữ liệu
- e. Ý kiến khác.

Câu 2. K là vùng tin khóa của tập tin dữ liệu Nhân_viên. Việc tìm kiếm các bản ghi của tập tin Nhân_viên với điều kiện tìm kiếm trên K sẽ hiệu quả khi:

- a. Các bản ghi được sắp thứ tự theo giá trị của K và không giới hạn điều kiện tìm kiếm trên K.
- b. Các bản ghi được sắp thứ tự theo giá trị của vùng tin K và điều kiện tìm kiếm trên K là “=”.
- c. Vùng tin K là vùng tin băm và điều kiện tìm kiếm trên K là “>”.
- d. Tồn tại chỉ mục trên vùng tin K và điều kiện tìm kiếm trên K là “<>”.
- e. Ý kiến khác.

Câu 3. Thao tác thêm bản ghi mới vào tập tin dữ liệu sẽ hiệu quả khi:

- a. Tập tin dữ liệu dạng không thứ tự (*unordered*)
- b. Tập tin dữ liệu dạng có thứ tự (*ordered*)
- c. Tập tin băm (*hashed*)
- d. Tập tin chỉ mục (*indexed*)
- e. Ý kiến khác

Câu 4. Tập tin tĩnh (*static*) và tập tin động (*dynamic*) được phân biệt dựa trên đặc điểm gì?

- a. Tần suất tái cấu trúc (*reorganize*) tập tin
- b. Tần suất truy vấn (*query*) nội dung của tập tin
- c. Tần suất sao lưu (*backup*) nội dung của tập tin
- d. Tần suất cập nhật (*update*) nội dung của tập tin
- e. Ý kiến khác.

Họ - Tên:

Mã Số Sinh Viên:

Câu 5. Xác định chi phí truy đạt khối trung bình trên tập tin dữ liệu băm (*hashed file*) khi điều kiện tìm kiếm trên vùng tin băm (*hashing field*) là “=”. Cho biết b là số khối dữ liệu hiện có trong tập tin.

- a. $O(1)$ b. $O(\log_2 b)$ c. $O(\ln b)$ d. $O(b)$
e. Ý kiến khác.

Câu 6. Cây chỉ mục B+ (*B+-tree*) có ưu điểm gì sau đây so với cây chỉ mục B (*B-tree*)?

- a. B+-tree chính là B-tree với các nút lá có thêm các con trỏ liên kết giữa chúng.
b. B+-tree chính là B-tree nhưng được dùng riêng cho chỉ mục thứ cấp.
c. B-tree thường có số nút và số mức nhiều hơn B+-tree nên chi phí hiệu chỉnh cân bằng cho B-tree thường cao hơn so với chi phí hiệu chỉnh cân bằng cho B+-tree.
d. Nút nội của B+-tree không có chứa con trỏ dữ liệu nên có khả năng chứa nhiều giá trị tìm kiếm hơn nút nội của B-tree.
e. Ý kiến khác.

Câu 7. Dạng chỉ mục nào sau đây có thể là chỉ mục dày (*dense index*)?

- a. Sơ cấp (*primary*)
b. Cụm (*clustering*)
c. Thứ cấp (*secondary*)
d. Tùy vào đặc điểm tập tin dữ liệu được chỉ mục, bất kỳ dạng chỉ mục nào (sơ cấp, cụm, thứ cấp) đều có thể là chỉ mục dày.
e. Ý kiến khác.

Câu 8. Giả sử vùng tính khóa V (*search key field*) chiếm 8 byte, kích thước *block* B là 1024 byte, *record pointer* Pr chiếm 7 byte, và *block pointer* P chiếm 6 byte. Xác định bậc của cây chỉ mục B. Biết rằng mỗi nút của cây chỉ mục B được chứa trọn trong 1 *block*.

- a. $p = 23$
b. $p = 25$
c. $p = 49$
d. Thông tin mô tả về cây chỉ mục B chưa đầy đủ nên không thể xác định được bậc của cây này.
e. Ý kiến khác.

Câu 9. Trong hệ quản trị cơ sở dữ liệu Oracle, người sử dụng có thể sử dụng hints (*/*+ */*) để dẫn hướng cho bộ tối ưu hóa truy vấn (*query optimization*) trong quá trình xử lý truy vấn. Khi đó, Oracle giả định điều gì sau đây?

- a. Người sử dụng có nhiều thông tin hơn về tổ chức dữ liệu (*data organization*) và các phương thức truy đạt (*access method*) trên dữ liệu của họ so với bộ tối ưu hóa truy vấn của Oracle.
b. Bộ tối ưu hóa truy vấn của Oracle không có đủ thông tin về các phương thức truy đạt trên dữ liệu của người sử dụng.

Họ - Tên:

Mã Số Sinh Viên:

c. Quá trình xử lý truy vấn của Oracle được chuyển sang chế độ do người dùng chỉ định (*user-specified mode*).

d. Câu truy vấn có hints chưa được định nghĩa trước trên dữ liệu của người sử dụng.

e. Ý kiến khác.

Câu 10. Chi phí giao tiếp (*communication cost*) nên được xem xét cho trường hợp xử lý truy vấn trên cơ sở dữ liệu nào sau đây?

a. Cơ sở dữ liệu lớn (*large database*)

c. Cơ sở dữ liệu phân tán (*distributed database*)

b. Cơ sở dữ liệu nhỏ (*small database*)

d. Bất kỳ dạng cơ sở dữ liệu nào

e. Ý kiến khác.

Câu 11. Cho công thức tính chi phí cho phép toán chọn (*selection*) dùng chỉ mục sơ cấp với cấu trúc chỉ mục đa mức động B+-tree trong trường hợp điều kiện chọn là điều kiện so sánh (<, <=, >, >=) trên thuộc tính khóa được chỉ mục: $C = x + b/2$; trong đó, **x** là số mức của B+-tree, **b** là số khối trong tập tin dữ liệu. **Giả định** (*assumption*) gì được sử dụng trong công thức này?

a. Phân nửa số nút trong chỉ mục chứa các con trỏ dữ liệu chỉ đến các khối trong kết quả.

b. Phân nửa số khối chỉ mục ở tầng lá chính là số khối chứa các bản ghi trong kết quả.

c. Chỉ phân nửa số khối dữ liệu trong tập tin cần được kiểm tra điều kiện chọn.

d. Phân nửa số bản ghi tương ứng là phân nửa số khối trong tập tin thỏa điều kiện chọn.

e. Ý kiến khác.

Câu 12. Sắp thứ tự ngoại (*external sorting*) được sử dụng trong việc xử lý truy vấn nào sau đây?

a. Phép chọn (*selection, σ*)

b. Phép kết với phương pháp hai vòng lặp lồng (*nested loop join*)

c. Phép giao (*intersection, \cap*)

d. Phép kết với phương pháp một vòng lặp (*single loop join*)

e. Ý kiến khác.

Câu 13. Một giao tác có thể bị ngưng thực thi (*aborted*) vì:

a. Yêu cầu từ bộ điều khiển tương tranh (*concurrency controller*)

b. Yêu cầu từ người dùng

c. Yêu cầu từ hệ điều hành

d. Câu a, b, và c đều đúng.

e. Ý kiến khác.

Họ - Tên:

Mã Số Sinh Viên:

Câu 14. Trong môi trường đa người dùng, tính cách ly/đơn lập (*isolation*) cần được đảm bảo bởi phân hệ (*subsystem*) nào của hệ quản trị cơ sở dữ liệu (*database management system, DBMS*)?

- a. Tối ưu hóa truy vấn (*query optimizer*)
- b. Thi hành ràng buộc toàn vẹn (*integrity constraint enforcer*)
- c. Điều khiển tương tranh (*concurrency control*)
- d. Phục hồi (*recovery*)
- e. Ý kiến khác.

Câu 15. Khi một giao tác T đã *commit*, những hoạt động nào sau đây cần được thực hiện cho T?

- a. Những thao tác đọc/ghi dữ liệu của T được ghi nhận vào sổ ghi nhật ký (*log*) của hệ thống.
- b. Đánh dấu thông tin về *commit* của T vào *log* bằng mục tin [commit, T].
- c. Cả hai hoạt động ở (a) và (b) cần được thực hiện cho T.
- d. Không hoạt động nào ở (a) và (b) cần được thực hiện cho T.
- e. Ý kiến khác.

Câu 16. Trong quá trình phục hồi, *redo* (tái thực hiện) là thao tác cần được thực hiện cho các tác vụ của một giao tác có đặc điểm gì sau đây?

- a. Giao tác đã *commit*, nhưng những tác động của giao tác này lên cơ sở dữ liệu chưa được ghi nhận bền vững trong cơ sở dữ liệu.
- b. Giao tác đạt đến điểm *commit*, nhưng những tác động của giao tác này lên cơ sở dữ liệu chưa được ghi nhận bền vững trong sổ ghi nhật ký (*log*) của hệ thống.
- c. Giao tác đang thực hiện các thao tác đọc/ghi thì bị ngưng (*aborted*) đột ngột.
- d. Giao tác đang thực hiện các thao tác đọc/ghi nhưng bị yêu cầu quay lui (*rollback*).
- e. Ý kiến khác.

Câu 17. Khi nói cần thực hiện *undo* (tháo gỡ) cho thao tác $\text{write}(X, T, 10, 15)$ của giao tác T; trong đó, 10 là giá trị cũ, 15 là giá trị mới, hệ thống cần làm gì?

- a. Cập nhật lại giá trị X trong cơ sở dữ liệu để giá trị của X là 15.
- b. Cập nhật lại giá trị X trong sổ ghi nhật ký (*log*) của hệ thống để giá trị của X là 15.
- c. Cập nhật lại giá trị X trong sổ ghi nhật ký (*log*) của hệ thống để giá trị của X là 10.
- d. Cập nhật lại giá trị X trong cơ sở dữ liệu để giá trị của X là 10.
- e. Ý kiến khác.

Câu 18. *Deadlock* (khóa chết) KHÔNG xảy ra với phiên bản của kỹ thuật khóa hai pha (*two-phase locking*) nào sau đây?

- a. Khóa hai pha cơ bản (*basic two-phase locking*)
- b. Khóa hai pha bảo thủ (*conservative two-phase locking*)
- c. Khóa hai pha nghiêm cách (*strict two-phase locking*)
- d. Không có phiên bản nào của kỹ thuật khóa hai pha lại không có hiện tượng khóa chết.
- e. Ý kiến khác.

Môn: **Hệ Quản Trị Cơ Sở Dữ Liệu** (503004)

Học kỳ 1 - 2013-2014

Ngày thi: 17/12/2013

Phòng thi:

Mã đề: **01**

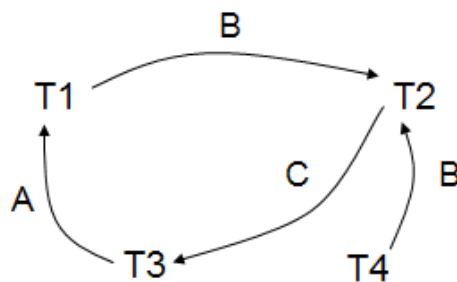
Họ - Tên:

Mã Số Sinh Viên:

Câu 19. Cho lịch biểu $S_{19} = r_3(X); r_2(X); w_3(X); r_1(X); r_3(Y); w_1(X)$. Xác định lịch biểu tuần tự (*serial schedule*) tương đương xung đột với lịch biểu S_{19} này.

- a. T3; T2; T1.
- b. T2; T3; T1.
- c. T1; T2; T3.
- d. Không tồn tại lịch biểu tuần tự tương đương xung đột với lịch biểu S_{19} này do lịch biểu này không khả tuần tự hóa xung đột.
- e. Ý kiến khác.

Câu 20. Cho đồ thị đợi (*wait-for graph*) của một lịch biểu dưới đây; trong đó, điều khiển tranh của các giao tác được thực hiện với kỹ thuật khóa hai pha. **Khóa chết (deadlock)** có xảy ra với lịch biểu này không?



- a. Không đủ thông tin về các giao tác để xác định liệu khóa chết có xảy ra hay không.
- b. Chỉ khi các giao tác đạt đến điểm *commit*, đồ thị đợi mới có đầy đủ thông tin và khi đó, việc xác định liệu khóa chết có xảy ra hay không mới được xác định đúng.
- c. Có
- d. Không
- e. Ý kiến khác.

Câu 21. Cho ma trận xung đột (*conflict matrix*) của kỹ thuật khóa 2 pha dùng `read_lock`, `write_lock`, và `unlock` như hình bên.

	Read	Write
Read	Y	N
Write	N	N

`write_lock` được thực hiện trong chế độ loại trừ (*exclusive mode*); có nghĩa là:

- a. Nếu giao tác T1 đang giữ khóa `write_lock (X)` trên X thì hệ thống sẽ không cấp bất kỳ khóa nào nữa cho các giao tác khác đang có yêu cầu khóa trên X.
- b. Nếu giao tác T1 đang giữ khóa `write_lock (X)` trên X thì hệ thống sẽ không cấp khóa ghi `write_lock` nữa cho các giao tác khác đang có yêu cầu khóa trên X.
- c. Nếu giao tác T1 đang giữ khóa `write_lock (X)` trên X thì hệ thống sẽ không cấp khóa đọc `read_lock` nữa cho các giao tác khác đang có yêu cầu khóa trên X.
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Họ - Tên:

Mã Số Sinh Viên:

Câu 22. Cho các giao tác T1 và T2 thực thi đồng thời như sau.

T1	T2
$r_1(X)$	
$w_1(X)$	
	$r_2(X)$
	$w_2(X)$
$r_1(Y)$	
(aborted)	

Sau khi đọc Y, T1 bị ngưng thực thi. Xác định tình trạng của T2 để đảm bảo việc thực thi của các giao tác này là đúng.

- a. T2 nên *commit* do T2 đã thực hiện xong các thao tác của T2.
- b. T2 nên bị ngưng thực thi và quay lui do T2 đã sử dụng giá trị X vừa mới được cập nhật bởi giao tác T1 bị ngưng thực thi.
- c. T2 nên tiếp tục thực hiện các thao tác còn lại của T2.
- d. Câu a, b, và c đều sai.
- e. Ý kiến khác.

Câu 23. Kỹ thuật khóa hai pha bảo thủ (*conservative two-phase locking*) nên được sử dụng trong hệ thống có đặc điểm gì sau đây?

- a. Thường có các giao tác có nhiều thao tác trên nhiều đối tượng dữ liệu và số lượng giao tác thực thi đồng thời ít.
- b. Thường có các giao tác có nhiều thao tác trên nhiều đối tượng dữ liệu và số lượng giao tác thực thi đồng thời nhiều.
- c. Thường có các giao tác có ít thao tác trên một số đối tượng dữ liệu và số lượng giao tác thực thi đồng thời ít.
- d. Thường có các giao tác có ít thao tác trên một số đối tượng dữ liệu và số lượng giao tác thực thi đồng thời nhiều.
- e. Ý kiến khác.

Câu 24. Việc duy trì nhiều phiên bản khác nhau của các đối tượng dữ liệu trong một số kỹ thuật điều khiển tương tranh nhằm mục đích gì sau đây?

- a. Lịch sử của các thao tác cập nhật được ghi nhận.
- b. Lịch sử thực thi của các giao tác được ghi nhận.
- c. Các thao tác đọc của các giao tác luôn được thực hiện với các phiên bản thích hợp.
- d. Các thao tác ghi của các giao tác luôn được thực hiện với các phiên bản thích hợp.
- e. Ý kiến khác.

Họ - Tên:

Mã Số Sinh Viên:

Câu 25. Cho nội dung sổ ghi nhật ký (*log*) của hệ thống như sau. Hệ thống gặp sự cố ngay khi giao tác T2 thực hiện thao tác đọc dữ liệu A. Bộ phục hồi (*recovery subsystem*) cần thực hiện các hoạt động gì để đảm bảo đặc tính **ACID** của các giao tác T1 và T2?

[start_transaction, T1]
[read_item, T1, A]
[start_transaction, T2]
[write_item, T1, A, 20, 25]
[read_item, T2, B]
[commit, T1]
[write_item, T2, B, 10, 18]
system crash → [read_item, T2, A]

- a. Đối với giao tác T1, nếu các ảnh hưởng của các thao tác của T1 chưa được ghi nhận bền vững trên cơ sở dữ liệu thì bộ phục hồi sẽ tái thực hiện (*redo*) các thao tác của T1 theo thứ tự duyệt nội dung *log* từ trên xuống. Đối với giao tác T2, nếu các ảnh hưởng của các thao tác của T2 đã được ghi nhận bền vững trên cơ sở dữ liệu thì bộ phục hồi sẽ tháo gỡ (*undo*) các thao tác của T2 theo thứ tự duyệt nội dung *log* từ dưới lên.
- b. Đối với giao tác T1 và T2, nếu các ảnh hưởng của các thao tác của T1 và T2 chưa được ghi nhận bền vững trên cơ sở dữ liệu thì bộ phục hồi sẽ tái thực hiện (*redo*) các thao tác của T1 và T2 theo thứ tự duyệt nội dung *log* từ trên xuống.
- c. Đối với giao tác T1 và T2, nếu các ảnh hưởng của các thao tác của T1 và T2 đã được ghi nhận bền vững trên cơ sở dữ liệu thì bộ phục hồi sẽ tháo gỡ (*undo*) các thao tác của T1 và T2 theo thứ tự duyệt nội dung *log* từ dưới lên.
- d. Đối với giao tác T1, nếu các ảnh hưởng của các thao tác của T1 đã được ghi nhận bền vững trên cơ sở dữ liệu thì bộ phục hồi sẽ tháo gỡ (*undo*) các thao tác của T1 theo thứ tự duyệt nội dung *log* từ dưới lên. Đối với giao tác T2, nếu các ảnh hưởng của các thao tác của T2 chưa được ghi nhận bền vững trên cơ sở dữ liệu thì bộ phục hồi sẽ tái thực hiện (*redo*) các thao tác của T2 theo thứ tự duyệt nội dung *log* từ trên xuống.
- e. Ý kiến khác.

Họ - Tên:

Mã Số Sinh Viên:

II. PHẦN TỰ LUẬN

Sinh viên trả lời phần tự luận vào **giấy làm bài**. Cho mỗi câu hỏi, sinh viên cần trình bày các bước thực hiện và có phần giải thích để có kết quả trả lời sau cùng.

Câu 26. Cho các giá trị phân biệt của vùng tin Mã_Phòng_Ban của tập tin Nhân_Viên như sau: 3, 5, 8, 2, 10, 14, 1. Một chỉ mục thứ cấp được định nghĩa trên vùng tin này.

- Trình bày các bước xây dựng chỉ mục thứ cấp này với cấu trúc chỉ mục B+-tree với bậc nút nội $p = 3$ và bậc nút lá $p_{leaf} = 2$. Giả sử các giá trị được đưa vào chỉ mục theo thứ tự đã cho từ trái sang phải. (1 điểm)
- Theo phương pháp tối ưu hóa dựa trên chi phí (*cost-based optimization*), xác định phương án thực thi tối ưu nhất có thể và chi phí tương ứng của phương án để lấy ra thông tin của tất cả nhân viên ở phòng ban 10 theo biểu thức truy vấn sau: $\sigma_{Mã_Phòng_Ban=10}(Nhân_Viên)$. Biết rằng tập tin Nhân_Viên có 300 bản ghi tương ứng ($r_{Nhân_Viên} = 300 records$) với 300 nhân viên và có hệ số phân khối là $bfr = 5$ bản ghi/khối. (1 điểm)

Câu 27. Cho lịch biểu $S_{27} = r_1(X), r_2(X), w_1(X), r_1(Y), w_2(X), c_2, w_1(Y), c_1$.

- Xác định đặc điểm khả phục hồi (*non-recoverable, recoverable, cascadeless, hay là strict*) của S_{27} . Giải thích lựa chọn của em. (1 điểm)
- Xác định đặc điểm khả tuần tự hóa xung đột (*serializable hay là non-serializable*) của S_{27} . Giải thích lựa chọn của em. (1 điểm)

Câu 28. Cho 2 giao tác T_1 và T_2 thực thi đồng thời như sau:

T_1	T_2
$r_1(X)$	$r_2(Z)$
$r_1(Z)$	$r_2(Y)$
$w_1(X)$	$w_2(Z)$
$r_1(Y)$	$r_2(X)$
$w_1(Z)$	

- Xác định lịch biểu khả tuần tự hóa xung đột được điều khiển tương tranh bằng kỹ thuật khóa 2 pha với bộ ba tác vụ khóa: *read_lock, write_lock, và unlock*. Lưu ý: mỗi giao tác trong lịch biểu đạt được phải được gán khóa theo đúng qui tắc của kỹ thuật khóa 2 pha. (1 điểm)
- Trong lịch biểu mà em đã thực hiện điều khiển tương tranh bằng kỹ thuật khóa 2 pha, khóa chết (*deadlock*) có xảy ra không? Giải thích tại sao khóa chết (*không*) xảy ra. (1 điểm)